Model Archive Summary for Suspended-Sediment Concentration at Station 11455167 Prospect Slough at Toe Drain Nr Courtland, Ca

This model archive summary summarizes the suspended-sediment concentration (SSC) model developed to compute 15-minute SSC timeseries for the period of record 10/28/2014 to 5/23/2018. This is the first suspended-sediment model developed for the site. The methods used follow USGS guidance as referenced in relevant Office of Surface Water/Office of Water Quality Technical Memorandum 2016.10 (USGS, 2016) and USGS Techniques and Methods, book 3, chap C4 (Rasmussen and others, 2009). This summary and model archive are in accordance with Attachment A of Office of Water Quality Technical Memorandum 2015.01 (USGS 2014).

Site and Model Information

Site number: 11455167
Site name: Prospect Slough at Toe Drain Nr Courtland, California
Location: Lat 38°17'25.39", long 121°39'44.01" referenced to North American Datum of 1983, CA, Hydrologic Unit 18020163
Equipment: A YSI EXO2 sonde began logging turbidity on October 28, 2014 and was removed from the station on May 23, 2018.

Model number: 11455167.SSC.WY2015.1
Model calibration data period: November 23, 2015 to May 3, 2018.
Model application date: October 28, 2014 to May 23, 2018.
Computed by: Tara Morgan-King, USGS, Sacramento, CA (tamorgan@usgs.gov)
Reviewed by: Anna Conlen, USGS, Sacramento, CA (aconlen@usgs.gov)

Physical Sampling Details and Sediment Data

All sediment data were collected using U.S. Geological Survey (USGS) protocols (USGS, 2006) and are stored in the National Water Information System (NWIS) database https://waterdata.usgs.gov/nwis. Discrete, boat-based samples are collected seasonally (roughly 6-12 times throughout the year) spanning the range of conditions targeting high flow and sediment concentration events as conditions permit. We collected an annual average of 8 samples/year including 13 observation samples in water year (WY) 2016, 8 in WY2017, and 2 in WY2018 before the sensor was removed.

Sample collection is consistent with approved field methods described in Edwards and Glysson (1999) and USGS (2006). The Equal Discharge Increment (EDI) method was used to determine the locations of five sampling verticals along the transect where discharge weighted suspended-sediment samples were collected. Each sampling vertical is located at the centroid of increments representing 20% of the total flow (5 verticals). Due to the tidal nature of the site, the EDI method was used to collect discharge-weighted samples because velocities are not always isokinetic (based on Table 4-5 from TWRI09A4, USGS 2006). A boat-based discharge measurement was collected immediately before sampling to determine the location of each vertical. A FISP US D-74 was used to collect the depth integrated samples at each vertical across the cross section. The maximum depth of the channel cross section can approach 9.5 feet but the average depth of the cross section is typically less than 7 feet. Sediment at this station is mostly fines (96% fines on average from sand/fine break analysis) and potential sampling bias due to non-isokinetic sampling is considered minimal. The site velocities during the deployment ranged from 2.6 ft/sec during flood tides to 6.2 ft/sec during ebb tide.

Samples were analyzed for SSC and sand/fine split at the USGS Santa Cruz, California sediment lab. Samples were analyzed for SSC by the filtration method. Each of the 5 verticals were analyzed individually at the lab for quality control purposes. Samples were not composited due to rapidly changing, tidal conditions. The set average SSC from 5 depth integrated verticals collected was computed and used in the calibration dataset. In rare occasions when the SSC at a vertical is deemed an outlier, a manual average was computed from fewer than 5 verticals and notes applied to the database. The samples collected on 9/27/2017 used the Equal Width Increment method to determine 10 sampling verticals along the cross section. All cross-sectional samples were included in the model as there were no outliers.

All sediment data were reviewed and marked as approved in the USGS NWIS Water-Quality System database (QWDATA) and made publicly available before being included in the calibration model. Approved samples were made publicly available on NWIS.

Surrogate Data

Continuous 15-minute turbidity data and discharge data were collected and computed by the USGS California Water Science Center and evaluated as possible explanatory variables for SSC. Data were measured using a YSI EXO2 sonde and reported in Formazin Nephelometric Turbidity Units (FNU). Data began logging on October 28, 2014 and the sonde was removed from the station on May 23, 2018. All surrogate turbidity data were computed, reviewed, and approved before using in the sediment calibration model per USGS guidelines (Wagner and others 2006). Discharge data were collected, computed, reviewed, and approved by the USGS California Water Science Center and retrieved from NWIS-TS. Methods to compute discharge follow Levesque and Oberg (2012). The 15-minute timeseries data are located at https://waterdata.usgs.gov/nwis/uv?site_no=11455167.

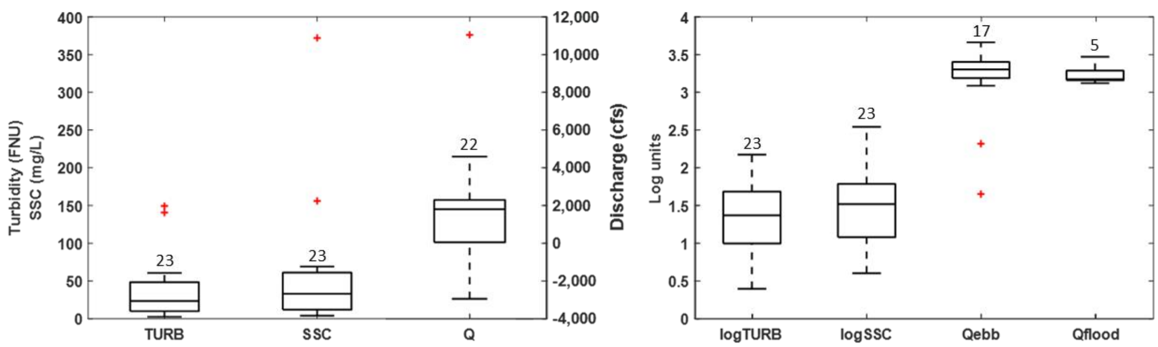
Model Calibration Dataset

The approved time-series turbidity data spanning the dates of the sediment constituent dataset were retrieved from NWIS-TS (Rasmussen and others 2009). The USGS Surrogate Analysis and Index Developer Tool (SAID) was used to pair the surrogate data with the discrete sediment data (Domanski and others 2015). Turbidity and discharge values were paired with each sediment sample observation from a matching max +/- of 15 minutes. The SAID manual is found at https://pubs.er.usgs.gov/publication/ofr20151177.

The regression model is based on 23 concurrent measurements of turbidity, SSC, and discharge. Summary statistics and the complete model-calibration data set are provided in the following sections.

Model Development

Multiple models were evaluated including simple linear regression (SLR) and multiple linear regression (MLR). The most common estimation technique is SLR, but MLR is an alternate tool for computing SSCs when the SLR *MSPE* statistic is larger than 20 percent (Rasmussen and others, 2009). The calibration dataset is composed of 23 concurrent turbidity, SSC, and discharge measurements (discharge n=22). Boxplots are shown below and note that due to negative tidal discharge values during the flood tide, ebb and flood values are shown separately with the absolute values during flood tides. USGS (2016) *recommends* a minimum of 36 paired observations, however the station was discontinued.



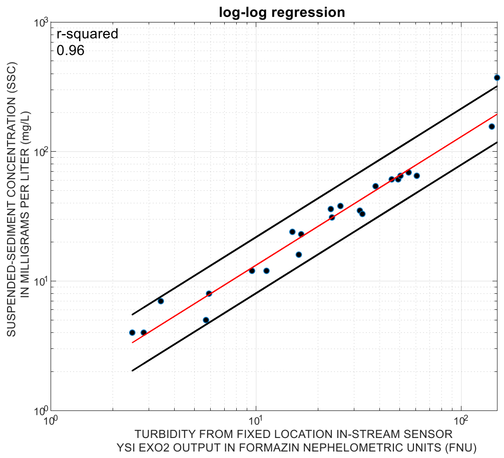
Model diagnostics and plots for model review were output using a variety of Matlab, SAID, and the R environment (R Core Team, 2018). The regression methods used are described in Helsel and Hirsch (2002). Table 3 in Rasmussen and others (2009) shows the best statistical diagnostics to help evaluate the models. The best model was chosen based on residual plots, model standard error, R^2, significance tests (p-values), correlation of explanatory variables, variance inflation factor (VIF), and PRESS (prediction error sum of squares) statistics. Values for the statistics and metrics were computed for various models and are included below along with all relevant sample data and more in-depth statistical information.

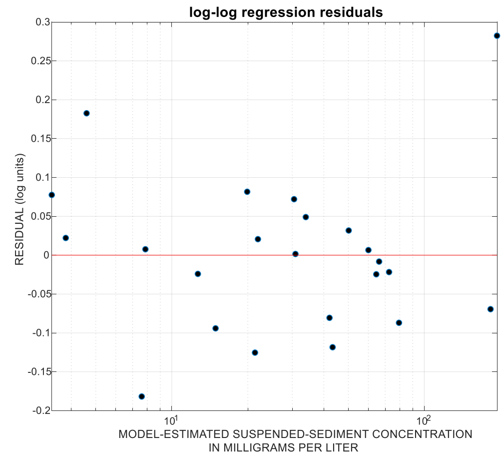
A variety of models were evaluated: Model 1) linear model with one explanatory variable (turbidity), Model 2) $\log_{10}$ transformed model with one explanatory variable (turbidity), Model 3) repeated medians method (Helsel and Hirsh, 2002) using one explanatory variable (turbidity), Model 4) linear model with two explanatory variables (turbidity and discharge), and Model 5) $\log_{10}$ transformed model with two explanatory variables (turbidity and discharge). Diagnostic statistics are summarized below for the five models evaluated. Discharge was not considered further as a second surrogate (in addition to turbidity) for an MLR model because 1) the discharge surrogate was not significant as a second variable to the model ($p>0.05$) and 2) the data gaps in the computed discharge time-series record dramatically decreased the total number of observations in the model and would also limit the resulting computed sediment time-series.

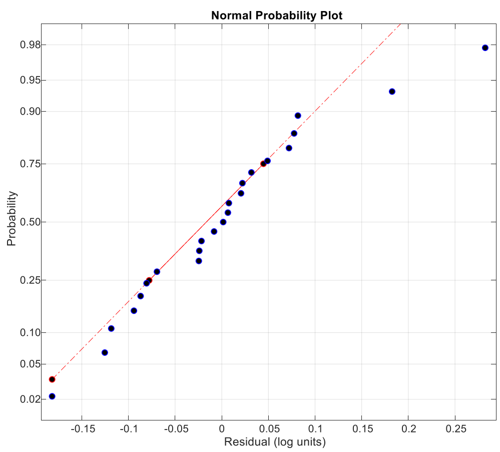
No.	R^2	R^2_a	RMSE	PRESS	MSPE	n	(type)
Model 1	0.82	0.81	33.5	63045	64.67	23	linear
Model 2	0.96	0.96	0.1	0.30	24.21	23	log
Model 3	0.74	0.73	40.7	106531	78.51	23	repeated median
Model 4	0.85	0.83	32.7	63639	63.17	22	multi-linear
Model 5	0.97	0.96	0.1	0.25	22.48	22	multi-log

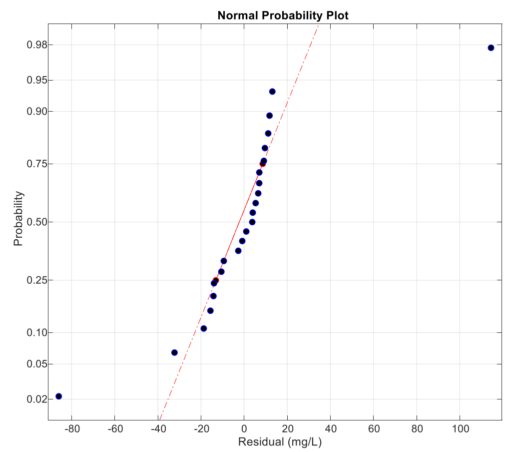
Flagged observations from the SAID outlier test criteria were evaluated. Standardized residuals from the models were inspected for values greater than 3 or less than negative 3. Values outside of the 3 to – 3 range are considered potential extreme outliers. The standardized residuals were reviewed from the SAID output reports and none of the samples were deemed as extreme outliers that should be removed from the model. All 23 observations were left in the model.

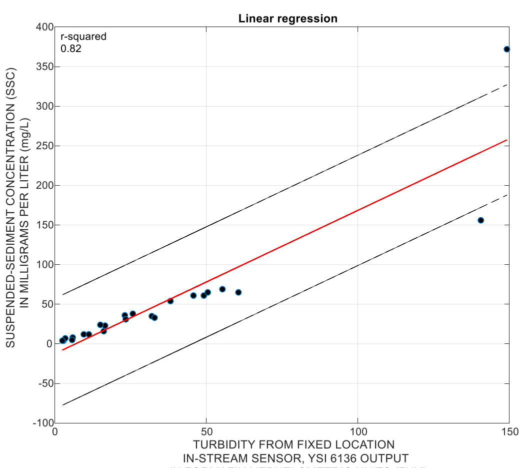
Of the models, the $\log_{10}$-transformed models had the highest R^2, lowest error and the residuals plots for the $\log_{10}$-transformed regression. Comparing the SLRs, Model 2 indicates a more homoscedastic pattern (constant variance) and a more normal distribution compared to the linear model (see the graphs below). Note that while the RMSE is not comparable between the log and linear models, we back transformed the predicted SSC and calculated the RMSE for the $\log_{10}$ model which was 24.3 (having a much lower error compared to the linear model – Model 1).

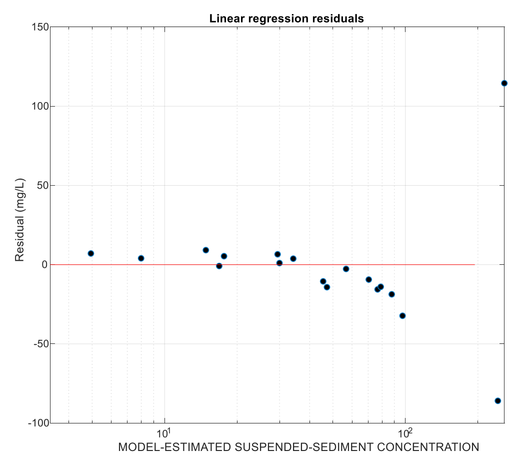

log-log regression
r-squared
0.96
SUSPENDED-SEDIMENT CONCENTRATION (SSC)
IN MILLIGRAMS PER LITER (mg/L)
TURBIDITY FROM FIXED LOCATION IN-STREAM SENSOR
YSI EXO2 OUTPUT IN FORMAZIN NEPHELOMETRIC UNITS (FNU)


log-log regression residuals
RESIDUAL (log units)
MODEL-ESTIMATED SUSPENDED-SEDIMENT CONCENTRATION
IN MILLIGRAMS PER LITER


Normal Probability Plot
Probability
Residual (log units)


Normal Probability Plot
Probability
Residual (mg/L)

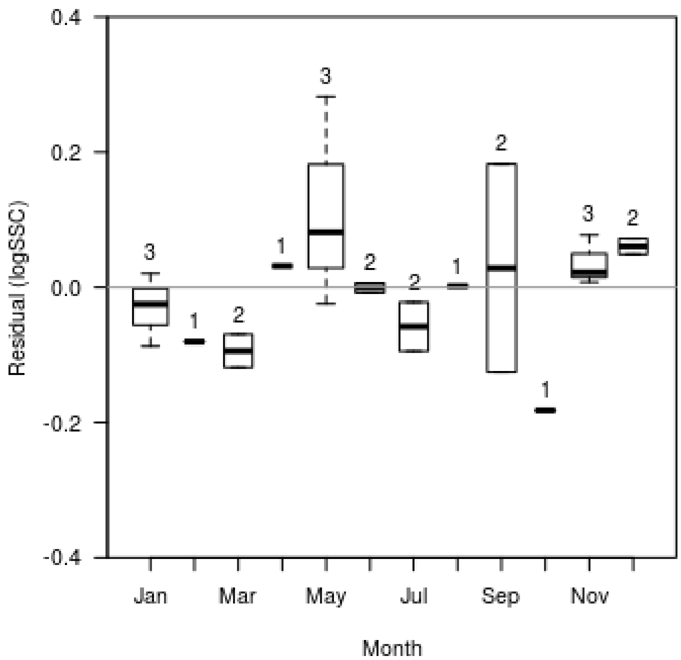

Linear regression
r-squared
0.82
SUSPENDED-SEDIMENT CONCENTRATION (SSC)
IN MILLIGRAMS PER LITER (mg/L)
TURBIDITY FROM FIXED LOCATION
IN-STREAM SENSOR, YSI 6136 OUTPUT
IN FORMAZIN NEPHELOMETRIC UNITS (FNU)

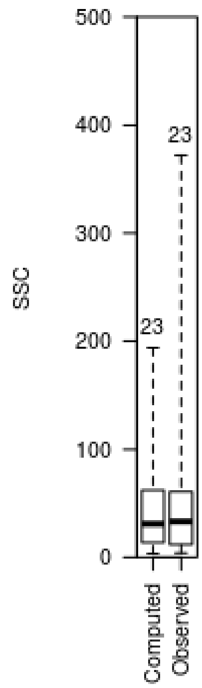

Linear regression residuals
Residual (mg/L)
MODEL-ESTIMATED SUSPENDED-SEDIMENT CONCENTRATION
IN MILLIGRAMS PER LITER

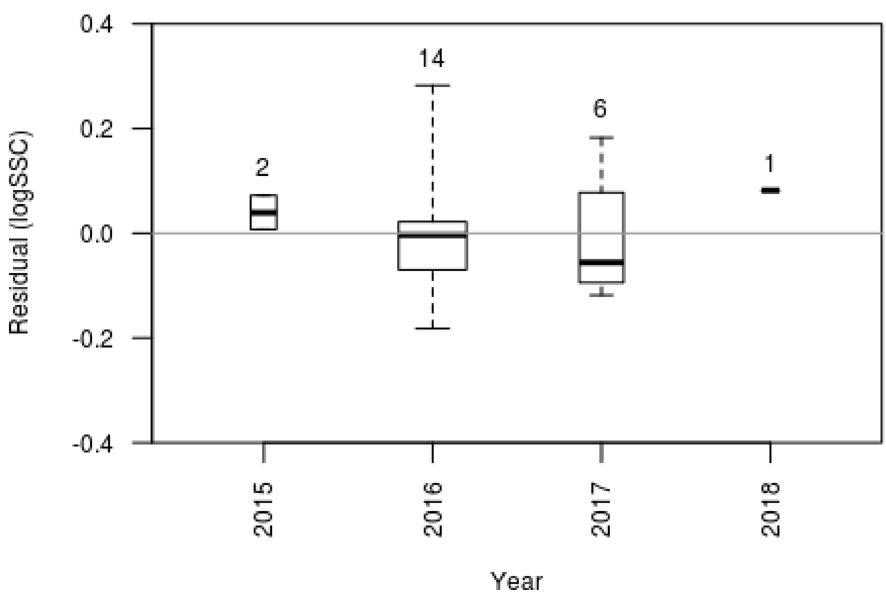
Plots of log_{10}SSC and explanatory variables and residual diagnostic plots

The following four plots were generated using a specialized R-Script developed by Patrick Eslick of the KSWSC and is located at the following address:

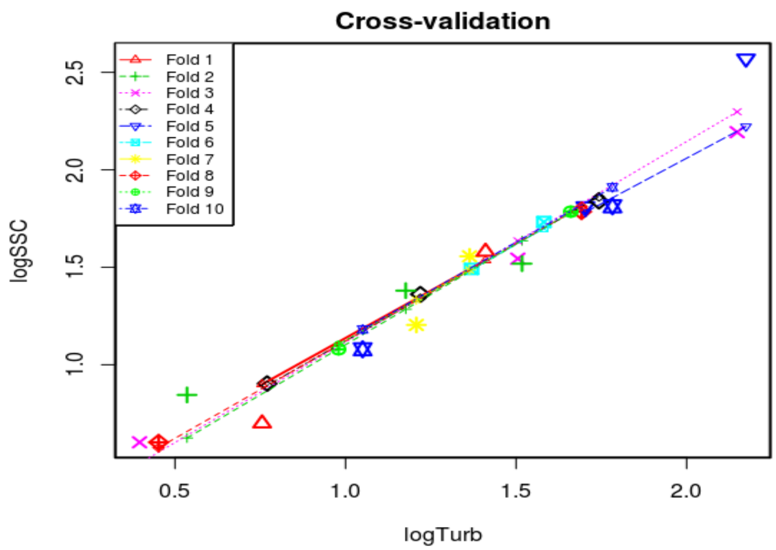
https://patrickeslick.github.io/ModelArchiveSummary/



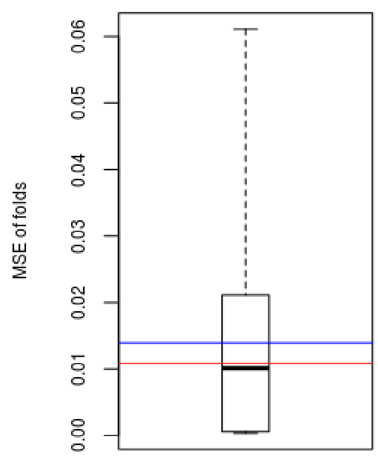




The graph below shows a k-fold cross validation with k=10 and the large points represent observations that were left out of each fold and are identified by the color and shape.




```
               Minimum MSE of folds:  0.000348
                  Mean MSE of folds:  0.013900
                Median MSE of folds:  0.010100
               Maximum MSE of folds:  0.061100
 (Mean MSE of folds) / (Model MSE):  1.280000
```




```
Red line - Model MSE

Blue line - Mean MSE of fold
```

Definitions
```
SSC: Suspended sediment concentration (SSC) in mg/l (80154)
Turb: Turbidity in FNU (63680)
```

MAS App Version 1.0

Model Summary

The final regression model for suspended-sediment concentration for site 11455167 is a simple log_{10}-transformed regression model based on 23 concurrent measurements of cross-sectional SSC samples and turbidity collected over approximately 3 water years from November 23, 2015 to May 3, 2018. The simple linear regression (SLR) model is shown below with basic model information, regression coefficients, correlation, summary statistics, and Duan's bias correction factor (Duan, 1983):

Linear Regression Model	Coefficient of Determination (R^2)
$log_{10}SSC = 0.132 + 0.992 * log_{10}Turb$	0.96

where

SSC is suspended-sediment concentration, in milligrams per liter, and
Turb is turbidity, in formazin nephelometric units, measured with a YSI EXO.

Because SSC was transformed during regression model development, the computed prediction may be biased and needs to be multiplied by a non-parametric smearing bias correction factor (BCF) which is shown below.

Model	Start date	End date	Linear Regression Model	*BCF*
1	10/28/2014	5/23/2018	$SSC = 10^{0.132} \times Turb^{0.992} \times BCF$	1.03

The log_{10}-transformed SLR model can be retransformed and corrected for bias resulting in the following equation:

$$SSC = 1.39Turb^{0.992}$$

Parameter	Minimum	Maximum
Turbidity (FNU) entire record	1.3	833
Computed SSC (mg/L)	5	*1097/410

*Extrapolation defined as computation beyond the range of the model calibration dataset may be used to extrapolate no more than 10 percent outside the range of the sample data used to fit the model. The original maximum computed SSC beyond the extrapolation limit is 1097 mg/L. Following USGS guidelines, the extrapolated, maximum computed SSC for this model is 410 mg/L.

Suspended-Sediment Concentration Record

The SSC record is computed using this regression model in the USGS National Real-Time Water Quality (NRTWQ) Web site. The complete record can be found at http://nrtwq.usgs.gov/ca.

Model

logSSC = + 0.992 * logTurb + 0.132

Variable Summary Statistics

	logSSC	SSC	logTurb	Turb
Minimum	0.602	4.0	0.396	2.49
1st Quartile	1.080	12.0	0.980	9.54
Median	1.520	33.0	1.370	23.40
Mean	1.430	51.8	1.310	35.50
3rd Quartile	1.790	61.0	1.690	49.20
Maximum	2.570	372.0	2.170	149.00

Basic Model Statistics

Number of Observations	23
Standard error (RMSE)	0.104
Average Model standard percentage error (MSPE)	24.2
Coefficient of determination (R^2)	0.959
Adjusted Coefficient of Determination (Adj. R^2)	0.957
Bias Correction Factor (BCF)	1.03

Explanatory Variables

	Coefficients	Standard Error	t value	Pr(>|t|)
(Intercept)	0.132	0.0625	2.1	4.76e-02
logTurb	0.992	0.0446	22.2	4.53e-16

Correlation Matrix

	Intercept	E.vars
Intercept	1.000	-0.938
E.vars	-0.938	1.000

Outlier Test Criteria

Leverage	Cook's D	DFFITS
0.261	0.193	0.590

Flagged Observations

Date	Time	logSSC	Estimate	Residual	Standard Residual	Studentized Residual	Leverage	Cook's D	DFFITS
5/17/2016	9:14	2.570	2.290	0.282	2.99	3.86	0.180	0.983	1.810
10/20/2016	10:34	0.699	0.881	-0.182	-1.84	-1.96	0.101	0.190	-0.657
9/27/2017	12:04	0.845	0.663	0.183	1.91	2.05	0.155	0.333	0.875

Model-Calibration Data Set

Observation Number	DateTime	logSSC	logTURB	SSC	TURB	Computed logSSC	Computed SSC	Residual	Normal Quantile	Censored Values
1	11/23/2015 9:50	0.90	0.77	8	5.9	0.90	8	0.01	0.22	--
2	12/17/2015 8:40	1.56	1.36	36	23.1	1.48	31	0.07	0.85	--
3	1/11/2016 9:56	1.36	1.22	23	16.6	1.34	23	0.02	0.33	--
4	1/26/2016 12:24	1.79	1.69	61	49.2	1.81	66	-0.02	-0.45	--
5	2/19/2016 9:17	1.54	1.51	35	32.0	1.62	43	-0.08	-0.70	--
6	3/15/2016 11:52	2.19	2.15	156	141	2.26	188	-0.07	-0.57	--
7	4/27/2016 9:07	1.73	1.58	54	38.2	1.70	52	0.03	0.57	--
8	5/17/2016 9:14	2.57	2.17	372	149	2.29	200	0.28	1.95	--
9	6/6/2016 11:41	1.79	1.66	61	45.8	1.78	62	0.01	0.11	--
10	6/8/2016 7:54	1.81	1.70	65	50.5	1.82	68	-0.01	-0.11	--
11	7/11/2016 8:35	1.84	1.74	69	55.3	1.86	75	-0.02	-0.22	--
12	8/9/2016 8:51	1.49	1.37	31	23.4	1.49	32	0.00	0.00	--
13	9/20/2016 14:54	1.20	1.21	16	16.1	1.33	22	-0.13	-1.48	--
14	10/20/2016 10:34	0.70	0.76	5	5.7	0.88	8	-0.18	-1.95	--
15	11/17/2016 10:24	0.60	0.45	4	2.8	0.58	4	0.02	0.45	--
16	12/17/2016 10:28	1.58	1.41	38	25.7	1.53	35	0.05	0.70	--
17	1/26/2017 10:11	1.81	1.78	65	60.6	1.90	82	-0.09	-0.85	--
18	3/8/2017 12:27	1.52	1.52	33	32.9	1.64	45	-0.12	-1.22	--
19	5/4/2017 11:46	1.08	0.98	12	9.5	1.10	13	-0.02	-0.33	--
20	7/13/2017 10:26	1.08	1.05	12	11.2	1.17	15	-0.09	-1.01	--
21	9/27/2017 12:04	0.85	0.54	7	3.4	0.66	5	0.18	1.48	--
22	11/8/2017 11:50	0.60	0.40	4	2.5	0.52	3	0.08	1.01	--
23	5/3/2018 11:55	1.38	1.18	24	15.0	1.30	20	0.08	1.22	--

References

Domanski, M.M., Straub, T.D., and Landers, M.N., 2015, Surrogate Analysis and Index Developer (SAID) tool (version 1.0, September 2015): U.S. Geological Survey Open-File Report 2015–1177, 38 p., http://doi.org/10.3133/20151177.

Duan, Naihua. 1983. Smearing estimate – A nonparametric retransformation method: Journal of the American Statistical Association. Volume 78-383. 605-610 p.

Edwards TK and Glysson GD. 1999. Field methods for measurement of fluvial sediment: U.S. Geological Survey Techniques of Water-Resources Investigations. Book 3, Chap. C2. 89 p. Available from: https://pubs.usgs.gov/twri/twri3-c2/pdf/TWRI_3-C2.pdf.

Helsel, D.R., and Hirsch, R.M., 2002, Statistical methods in water resources-Hydrologic analysis and interpretation: U.S. Geological Survey Techniques of Water-Resources investigations, book 4, chap. A3, 510 p.

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. Available from: https://www.R-project.org/.

Rasmussen P, Gray JR, Glysson GD, Ziegler AC. 2009. Guidelines and procedures for computing time-series suspended-sediment concentrations and loads from in-stream turbidity-sensor and streamflow data. Book 3 Applications of Hydraulics, Section C. 52 p. Available from: https://pubs.usgs.gov/tm/tm3c4/pdf/TM3C4.pdf.

[USGS] U.S. Geological Survey. 2006. National field manual for the collection of water quality data: U.S. Geological Survey Techniques of Water-Resources Investigations. Book 9, Chapter A4. Available from: https://pubs.usgs.gov/twri/twri9a4/twri9a4_Chap4_v2.pdf.

[USGS] U.S. Geological Survey, 2014, Policy and guidelines for archival of surface-water, groundwater, and water-quality model applications: Office of Groundwater Technical Memorandum 2015.02, Office of Surface Water Technical Memorandum 2015.01, Office of Water Quality Technical Memorandum 2015.01, Available from: https://water.usgs.gov/admin/memo/SW/sw2015.01.pdf

[USGS] U.S. Geological Survey. 2016. Policy and guidance for approval of surrogate regression models for computation of time series suspended-sediment concentrations and loads: Office of Surface Water Technical Memorandum 2016.07. Available from: https://water.usgs.gov/admin/memo/QW/qw2016.10.pdf.

Wagner RJ, Boulger RW, Jr, Oblinger CJ, Smith BA. 2006. Guidelines and standard procedures for continuous waterquality monitors: station operation, record computation, and data reporting: U.S. Geological Survey Techniques and Methods 1-D3. Available from: https://pubs.usgs.gov/tm/2006/tm1D3/pdf/TM1D3.pdf.